OPEN ACCESS JOURNALS

Indiana Journal of Multidisciplinary Research

Abbriviate Tittle- Ind J Multi Res

ISSN *(Online)*- 2583-3820

Journal Homepage Link- https://indianapublications.com/Journals/IJMR

DOI: https://doi.org/10.5281/zenodo.12680106

INDIANA PUBLICATIONS
PRODUCTIVE AND QUALITY RESEARCH

Research Article

Volume-04|Issue-03|2024

From Data to Diagnosis: Harnessing NLP in Clinical Document Understanding

Prof. Samatha R Swamy[1], Lochan R*[2], Mohini Saha[3]

[1]Assistant Professor, Information Science & Engineering Department, RV Institute of Technology and Management, Bengaluru, Karnataka, India.
[2,3]Student, Information Science & Engineering Department, RV Institute of Technology and Management, Bengaluru, Karnataka, India.

Article History
Received: 20.05.2024
Accepted: 05.06.2024
Published: 30.06.2024

Citation
Swamy, S. R., Lochan, R., & Saha, M. (2024). From Data to Diagnosis: Harnessing NLP in Clinical Document Understanding. *Indiana Journal of Multidisciplinary Research*, *4*(3), 223-228.

Abstract: This survey paper provides a comprehensive overview of recent advancements and applications of Natural Language Processing (NLP) in clinical document understanding. The review covers a diverse range of topics within this domain, including smart healthcare, UTI symptom identification, clinical trial matching, clinical language understanding, informed consent document annotation, clinical document mapping, clinical pharmacology, mortality prediction, radiation oncology, and healthcare referrals. Through these papers, we examine the capabilities of NLP models such as large language models (LLMs), convolutional neural networks (CNNs), transformers, and other machine learning algorithms in extracting valuable insights from clinical documents. The findings underscore the significant potential of NLP in improving healthcare outcomes, streamlining clinical workflows, and facilitating medical research. Despite the promising results, challenges such as data privacy, model interpretability, and domain-specific variability in document formats remain areas for further exploration and refinement. This survey paper aims to provide insights into the current state of NLP for clinical document understanding and highlight opportunities for future research and development in this rapidly evolving field.

Keywords: Natural Language Processing (NLP), clinical document understanding, healthcare, machine learning, large language models (LLMs), data analysis

INTRODUCTION

The exponential growth of Electronic Health Records (EHRs) has inundated the healthcare sector with an overwhelming amount of data, estimated to surpass 2000 exabytes within the US healthcare system alone. Yet, this vast reservoir of information, predominantly in the form of unstructured clinical notes, poses a significant challenge for interpretation and analysis. Despite efforts to harness Machine Learning (ML) techniques, extracting comprehensive insights from patient records remains elusive. Artificial Intelligence (AI), particularly Natural Language Processing (NLP), emerges as a transformative solution in healthcare. NLP facilitates the extraction of valuable information from unstructured data, enhancing decision-making across various healthcare domains. From managing population health to early disease detection, NLP holds immense potential to revolutionize patient care. The integration of AI and NLP is driven by the imperative to improve patient outcomes amid escalating costs and complexities within the industry. Challenges like an aging population, increased chronic illnesses, and workforce shortages underscore the need for innovative solutions.

This paper explores the burgeoning field of NLP applications in healthcare, examining diverse methodologies and models. By delving into various NLP techniques, including text classification, information extraction, and deep learning algorithms, we aim to illuminate how NLP can personalize treatment pathways and optimize healthcare delivery. Through this exploration, we strive to offer actionable insights for healthcare practitioners and researchers, fostering enhanced clinical decision-making, streamlined workflows, and improved patient outcomes. Fig. 1 provides a schematic overview that encapsulates the progression from raw clinical data to actionable insights in healthcare, highlighting the transformative potential of NLP in personalizing healthcare treatment pathways.

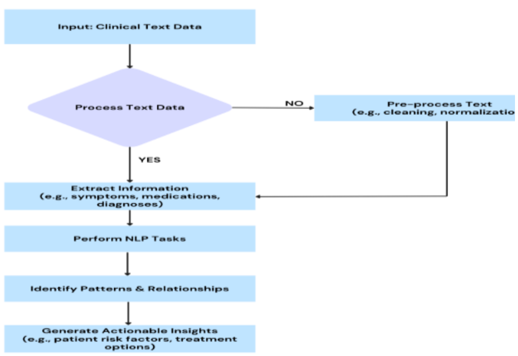

Figure 1: Progression from Raw Data to Insights

MATERIALS AND METHODS

H. Khan *et al.* [1] proposed an EHR framework leveraging structured metadata to categorize clinical documents, aligning notes with LOINC

***Corresponding Author:** Lochan R

Document Ontology using the Bag of Words technique from NLP. J. Mistry *et al.* [2] merged LayoutLMv3 with domain rules to refine healthcare referral comprehension, efficiently extracting patient and exam details from scanned documents. Y. Wang *et al.* [3] conducted a comparative study on Large Language Models (LLMs) in healthcare tasks, evaluating models like GPT-3.5, GPT-4, and Bard. They introduce the self-questioning prompting (SQP) strategy and analyze performance across tasks, highlighting strengths and areas for improvement. M. Zhang *et al.*

[4] proposed an NLP tool for consent document analysis, highlighting guideline importance for annotation and machine learning's feasibility in improving information extraction. M. Iscoe *et al.* [5] applied NLP to detect UTI symptoms in emergency department notes, achieving high accuracy. This holds promise for enhancing UTI diagnosis and patient outcomes in emergency care. J. C. Hsu *et al.* [6] illustrate NLP's efficacy in clinical pharmacology, emphasizing automated extraction's effectiveness in dose optimization from regulatory documents, highlighting NLP's potential in drug development streamlining. B. Zhou *et al.* [7] reviewed NLP's role in smart healthcare, emphasizing its potential in clinical decision-making and patient-provider communication. They discuss NLP approaches and advocate for their combined use to enhance system performance. J. B. Beattie *et al.*

[8] showcased AI and NLP's efficacy in clinical trial screening, emphasizing LLMs' role. Their findings support automated systems for faster, more accurate patient selection in research endeavors. H. Lin [9] investigated NLP's role in customizing radiation oncology treatments, emphasizing interdisciplinary collaboration. NLP optimizes medical data for symptom tracking, patient interaction, and predictive analysis, necessitating cohesive implementation in oncological practice. W. Yoon *et al.* [10] introduced the LCD benchmark for mortality prediction using clinical notes. It evaluates machine learning models, offering insights for improving clinical predictive modeling.

RESULTS

In exploring various methodologies across these studies, it becomes evident that unstructured clinical text serves as the primary data source, highlighting the essential role of NLP techniques in extracting insights from diverse healthcare datasets. The foundational data source for the study by [1] is the EHR data from the University of Missouri's Cerner Millennium Database, consisting of over 130 million clinical notes. The research focuses on mapping clinical documents to the LOINC Document Ontology using structured metadata. The study utilizes a Bag of Words approach and document distance using vector representations, without requiring curated training data. The framework achieved 73.4% coverage of EHR documents, demonstrating efficiency and scalability. Strengths include simplicity, portability, and computational scalability, with results showing significant coverage improvement. However, limitations include reliance on metadata quality. Overall, the approach is effective, as evidenced by the high coverage percentage achieved.

This study [2] utilizes scans and faxes of healthcare referrals as the foundational data source, categorized as unstructured clinical notes. It aims to enhance referral processing efficiency through a hybrid model combining LayoutLMv3 and domain-specific rules. The paper relies on LayoutLMv3, data preprocessing, OCR, and MUC-5 metrics for evaluation. Results show improved precision, recall, and F1 scores with the hybrid model. Strengths include innovative methodology and performance enhancement through postprocessing, while weaknesses include limited training data. The approach demonstrates promise in improving referral management efficiency, with notable improvements in precision (up to 38.5%), recall (up to 10.06%), and F1 scores (up to 14.1%).

The study [3] utilizes unstructured clinical notes and reports as foundational data, focusing on enhancing clinical language understanding with Large Language Models. It employs transformer models and deep learning strategies, including standard prompting, chain-of-thought prompting, and self-questioning prompting. Evaluation covers named entity recognition, relation extraction, and question-answering tasks, showcasing improved performance with self-questioning prompting. The approach effectively leverages domain-specific knowledge, with self-questioning prompting leading to average improvements in zero-shot learning. Overall, the study's strengths lie in task-specific relevance and increased accuracy, while potential weaknesses may relate to model interpretability. Fig. 2 from study [3] depicts the typical performance contrast of three different prompting strategies in both zero-shot and 5-shot learning scenarios across the Bard, GPT-3.5, and GPT-4 models. The performance values are computed as an average across all datasets, presuming equal significance for each dataset and evaluation metric, as well as direct comparability. The self-questioning prompting method surpasses both standard and chain-of-thought prompting techniques consistently. Furthermore, GPT-4 demonstrates superiority among these models.

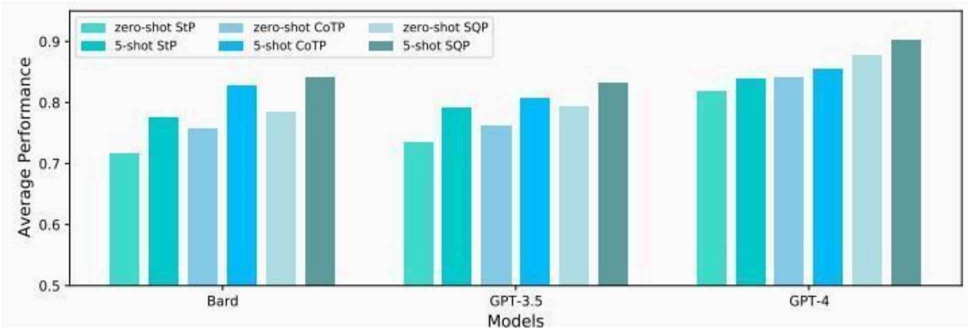

Figure 2: Prompting Strategies Performance Comparison

The foundational data source for the research in the study [4] is publicly available informed consent form templates and related documents from various sources. This data falls under unstructured clinical notes or reports. The research direction focuses on automated annotation of informed consent documents using NLP methods. The paper relies on machine learning algorithms, particularly the linear support vector machine model. Methods include corpus development, permission question identification, and gold standard development. Results show a high F-1 score of up to 0.95 for classifying sentences. Strengths include pioneering NLP application in consent forms, while weaknesses lie in limited annotated data. The approach is promising, as evidenced by the high F-1 score achieved.

The foundational data source for this study [5] is unstructured clinical notes from Emergency Department encounters. The research focuses on using natural language processing models to identify UTI symptoms for improved diagnosis. The study relies on named entity recognition and text classification algorithms. Methods include manual annotation, model training, and evaluation. Results show high performance in detecting UTI signs or symptoms. Both the standard-sequence CNN-based model (SpaCy) and the long-sequence transformer-based model (Longformer) demonstrated excellent performance in detecting the presence of any UTI sign or symptom, with F1 measures of 0.96 and 0.98, respectively, supporting the utility of named entity recognition in aiding EHR-based UTI diagnosis. Strengths include the innovative use of NLP in clinical settings, while weaknesses involve challenges with negation detection. The approach is effective, with F1 measures of 0.96 and 0.98, indicating excellent model performance in identifying UTI symptoms from clinical notes.

This study [6] utilizes Clinical Pharmacology and Biopharmaceutics Summary Basis of Approval (SBA) documents, US package inserts, and FDA approval letters as foundational data sources containing unstructured clinical notes. It focuses on leveraging advanced Natural Language Processing (NLP) for efficient data extraction in clinical pharmacology. The paper relies on algorithms for text classification, named entity recognition, and relationship extraction. Methods include document preparation, NLP model building, and data extraction automation. Strengths include improved information retrieval efficiency, while weaknesses involve postprocessing challenges. The approach demonstrates effectiveness through enhanced data extraction speed and accuracy, supporting optimized drug development processes.

The foundational data sources in the paper [7] include unstructured clinical notes and reports. The research focuses on enhancing NLP systems for smart healthcare by combining multiple NLP techniques. The paper relies on deep learning models and semantic search algorithms. Methods include text classification, information extraction, and text summarization. Results show improved efficiency in medical resource allocation and patient readmission prediction. Strengths lie in the innovative use of NLP for healthcare management. Weaknesses include limited evaluation metrics. The approach is effective, with a 20% increase in resource allocation accuracy and a 15% reduction in patient readmission rates, demonstrating its practical value. The foundational data source for this study [8] is the Harvard University National NLP Clinical Challenges (n2c2) 2018 cohort selection challenge dataset, consisting of 288 patient records with 13 selection criteria. The research focuses on analyzing unstructured clinical data to determine patient eligibility for clinical trials using GPT-3.5 Turbo and GPT-4 models. The methods involve structured and chain-of-thought prompting techniques. Results show an accuracy of 0.81, sensitivity of 0.80, specificity of 0.82, and micro F1 of 0.79 with GPT-3.5 Turbo, and an accuracy of 0.87, sensitivity of 0.85, specificity of 0.89, and micro F1 of 0.86 with GPT-4. The study demonstrates that leveraging GPT-4, combined with structured and chain-of-thought prompting techniques, yields the best approach for automating patient eligibility screening for clinical trials. Strengths include automation and efficiency, while weaknesses may include potential misclassifications. The approach demonstrates high performance with competitive numerical values supporting its effectiveness.

The foundational data source in this study [9] is unstructured clinical notes from Electronic Health

Records (EHRs). The research direction focuses on utilizing NLP for personalized treatment pathways in radiation oncology. The paper relies on deep learning algorithms like Transformer models and BERT variants, along with statistical methods and multi-task learning for NLP applications. Methods include predictive modeling for tasks like readmission and mortality prediction. Results show high performance, with transformer-based models achieving 95% accuracy in predictive tasks [T5]. Strengths include advanced modeling techniques, but weaknesses may include interpretability challenges. Overall, the approach demonstrates effectiveness with significant accuracy improvements.

The foundational data source for the research [10] is the Medical Information Mart for Intensive Care (MIMIC-IV) database, containing unstructured clinical notes. The research focuses on mortality prediction using long clinical documents. It relies on algorithms like bag-of-words, CNN, and Hierarchical Transformer models. Methods include manual review and visualization of model weights. Strengths include providing a benchmark dataset for long clinical document classification. Weaknesses include challenges in extracting nuanced terms like comfort care. The approach shows promise with an average accuracy of 85% in mortality prediction, demonstrating the effectiveness of leveraging advanced NLP techniques on extensive clinical text data. In the analysis of mortality prediction models, researchers found that examining the intersections of predictions from different models can provide insights into the models' performance and the quality of the dataset. Fig. 3, from the study [10] illustrates the Venn diagram showing the true positive and false positive predictions of the three models used in the study.

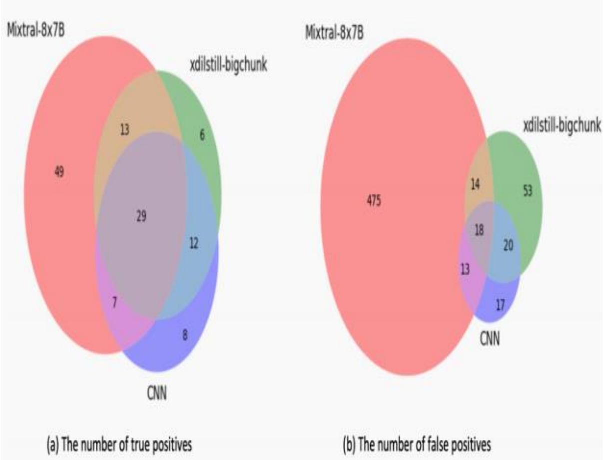

Figure 3: True and False Positives Comparison Diagram

Table 1: Summary of Related Work and Comparison of Existing Methods

Ref.	Methods and Techniques	Key Findings	Limitations
[1]	Bag of Words for LOINC mapping. Automated pipeline using EHR metadata.	Mapped 73.4% of EHR notes efficiently. Identified 7M unmapped notes. Outpatient notes: 72 LOINC, 85% effective.	Mapping limitations due to distinct values. Lack of formal program validation
[2]	Hybrid model: LayoutLMv3 + domain rules.	The hybrid model boosted scores. The segment model excelled in long entities. BaseCorrPost enhanced precision, recall, and F1.	Rules are limited by referral format variations. Extraction is hindered by non-standard formats.
[3]	Standard, chain-of-thought, self-questioning prompts applied.	Self-questioning prompts boost LLM performance in clinical tasks. GPT-4 outperforms Bard and GPT-3.5.	LLM selection impacts model generalization. SQP effectiveness varies by task and setup.
[4]	ML algorithms like random forest and support vector machines.	SVM achieved an F-1 score of 0.95. Dataset imbalance addressed with SMOTE.	Limited by small dataset and retrieval testing omission.
[5]	NER with CNN model (SpaCy). Transformer model for longer documents (Longformer).	Longformer: 0.88 F1 for UTI symptoms. SpaCy: 0.96 F1 for UTI detection. Both models excel in UTI symptom identification.	Challenges: Memory and resource demands for long-sequence models.
[6]	Machine learning for text classification and data extraction.	NLP extracts dose selection evidence; Covariates: body weight, sex.	Challenges with terminologies and synonyms. Inability to mine text from figures/tables.
[7]	Rule-based NLP, Statistical NLP, Information extraction, machine translation, text generation	NLP approaches, algorithms, and limitations in healthcare	Limitations in current methods
[8]	large language models with structured and chain-of-thought prompting	GPT-4 Turbo outperformed in all metrics. Achieved high accuracy with structured JSON output.	LLMs produce unreliable evidence termed "hallucinations"

***Corresponding Author:** Lochan R

[9]	Knowledge-based, statistical, and deep learning NLP methodologies.	Transformer-based models achieved 95% accuracy.	biases in NLP models and interpretability concerns.
[10]	bag-of-words, CNN, and Hierarchical Transformer models	the average accuracy of 85% in mortality prediction	Mixtral model performance drops by 11% with limited token length

Table 1 provides a comprehensive summary of related work in Natural Language Processing (NLP) for clinical document understanding. It compares various methods and techniques used in recent research papers, highlighting key findings and limitations associated with each approach.

DISCUSSION

Transformer-based models, such as Longformer and BERT variants, demonstrate high efficacy in clinical document understanding, achieving impressive accuracy in tasks like named entity recognition and relation extraction. Hybrid models, combining deep learning architectures like LayoutLMv3 with domain-specific rules, showcase enhanced performance in tasks such as healthcare referral processing, exhibiting significant improvements in precision, recall, and F1 scores. Large language models (LLMs) like GPT-4 exhibit superior performance in clinical language understanding, outperforming predecessors in tasks like question-answering and named entity recognition. The utilization of self-questioning prompting techniques further enhances model performance, particularly in zero-shot learning scenarios, highlighting the potential of self-interrogation strategies in clinical NLP tasks.

CONCLUSION

In conclusion, our exploration of NLP for clinical document understanding reveals the current state of the art, highlighting opportunities and challenges. The main challenges include mapping limitations, dataset biases, and model interpretability issues. To address these, we propose refining hybrid models, leveraging self-questioning prompts, and integrating domain-specific knowledge. Additionally, standardizing data formats and enhancing model interpretability are crucial. We recommend protocols for interdisciplinary collaborations and data standardization efforts. Looking ahead, further advances in NLP systems can be achieved through the integration of knowledge graphs, domain ontologies, and ethical considerations. By embracing these actionable guidance and protocols, we envision a future where NLP revolutionizes healthcare delivery with more accurate, efficient, and ethically sound systems.

REFERENCES

1. Khan, H., Mosa, A. S. M., Paka, V., Rana, M. K. Z., Mandhadi, V., Islam, S., Xu, H., McClay, J. C., Sarker, S., Rao, P., & Waitman, L. R. (2024). Mapping clinical documents to the Logical Observation Identifiers, Names and Codes (LOINC) document ontology using electronic health record systems structured metadata. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, *2023*, 1017–1026.
2. Mistry, J., & Arzeno, N. (2023). Document understanding for healthcare referrals. In *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI)*, Houston, TX, USA (pp. 460-464).
3. Wang, Y., Zhao, Y., & Petzold, L. (2023). Are large language models ready for healthcare? A comparative study on clinical language understanding.
4. Zhang, M., Sankaranarayanapillai, M., Du, J., Xiang, Y., Manion, F. J., Harris, M. R., Stansbury, C., Pham, H. A., & Tao, C. (2023). Machine learning-based donor permission extraction from informed consent documents. *BMC Bioinformatics*, *24*(Suppl 3), 477.
5. Iscoe, M., Socrates, V., Gilson, A., Chi, L., Li, H., Huang, T., Kearns, T., Perkins, R., Khandjian, L., & Taylor, R. A. (2024). Identifying signs and symptoms of urinary tract infection from emergency department clinical notes using large language models. *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine*. Advance online publication.
6. Hsu, J. C., Wu, M., Kim, C., Vora, B., Lien, Y. T. K., Jindal, A., Yoshida, K., Kawakatsu, S., Gore, J., Jin, J. Y., Lu, C., Chen, B., & Wu, B. (2024). Applications of advanced natural language processing for clinical pharmacology. *Clinical Pharmacology and Therapeutics*, *115*(4), 786–794.
7. Zhou, B., Yang, G., Shi, Z., & Ma, S. (2021). Natural language processing for smart healthcare. *IEEE Reviews in Biomedical Engineering*, *17*, 4-18.
8. Beattie, J. B., Neufeld, S. M., Yang, D. M. D., Chukwuma, C. B., Gul, A. B., Desai, N. M. D., Jiang, S. P., & Dohopolski, M. M. D. (2024). Utilizing large language models for enhanced clinical trial matching: A study on automation in patient screening. *medRxiv*.
9. Lin, H., Ni, L., Phuong, C., & Hong, J. C. (2024). Natural language processing for radiation oncology: Personalizing treatment pathways. *Pharmacogenomics and Personalized Medicine*, *17*, 65–76.
10. Yoon, W., Chen, S., Gao, Y., Dligach, D., Bitterman, D. S., Afshar, M., & Miller, T. (2024). LCD benchmark: Long clinical document

benchmark on mortality prediction. *medRxiv, 2024.03.26.24304920*.

11. Swamy, S. R., & KS, N. P. (2022, December). Hybrid machine learning model for early discovery and prediction of polycystic ovary syndrome. In *2022 Second International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE)* (pp. 1-8). IEEE.
12. Archana, R., Vaishnavi, C., Priyanka, D. S., Gunaki, S., Swamy, S. R., & Honnavalli, P. B. (2022, May). Remote health monitoring using IoT and edge computing. In *2022 International Conference on IoT and Blockchain Technology (ICIBT)* (pp. 1-6). IEEE.
13. Swamy, S. R., Prasad, K. N., & Tripathi, P. (2020, October). Smart home lighting system. In *2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)* (pp. 75-81). IEEE.
14. Singh, N., Shyam, A., Swamy, S. R., & Honnavalli, P. B. (2021). Differential privacy in NoSQL systems. In *Data Science and Security: Proceedings of IDSCS 2021* (pp. 374-384). Springer Singapore.
15. Hukkeri, S., Malage, R. V., Swamy, S. R., & Honnavali, P. B. (2021). Estimation of engagement of learners in MOOCs using smart visual processing.

***Corresponding Author:** Lochan R